The multi-FRI Protocol

Angus Gruen, Dmitry Vagner

1 Preliminaries

1.1 Notational Conventions

We denote an interval beginning with 0 as $\underline{n} = [0, n-1]$. Given a range $J = [m, n]$, it will also be useful for us to define that range, but excluding the first term, by $J^* = [m+1, n]$.

1.2 Probability, Distance, and Polynomials

We denote the probability of an event E by $\mathbb{P}[E]$. In the case that event $E(x)$ depends on a random variable $x \in U$, we denote the probability as $\mathbb{P}_{x \in U}[E(x)]$, or simply $\mathbb{P}_x[E(x)]$ if U is apparent from context. Any variables appearing in the expression of E but not denoted in the subscript of $\mathbb{P}$ are assumed to be constant. In much of what follows, we use probabilistic reasoning in the sense that the implication $E \implies E'$ entails the probability bound $\mathbb{P}[E] \leq \mathbb{P}[E']$.

Consider the functions $f, g : A \to B$, where A is a finite set. We define the *Relative Hamming Distance* between them as

$$d(f, g) = \frac{|\{x \in A \mid f(x) \neq g(x)\}|}{|A|} \tag{1}$$

Equivalently, this can be represented as a probability

$$d(f, g) = \mathbb{P}_{x \in A}[f(x) \neq g(x)], \tag{2}$$

One can think of this distance as the ratio of evaluations of f that need to be changed in order for it to equal g. This interpretation makes apparent the triangle inequality $d(f, h) \leq d(f, g) + d(g, h)$, since we can change f to h by first changing f to g and then changing g to h. Since it is clear that $d(f, g) = d(g, f)$ and that $d(f, g) = 0$ if and only if $f = g$, we have that the Relative Hamming distance gives the set of functions $A \to B$ the structure of a metric space. As is typical of metric spaces (M, d), we can define the distance between an element $f \in M$ and a subset $S \subset M$ as follows.

$$d(f, S) = \min_{g \in S} d(f, g) \tag{3}$$

One extension of distance that will be useful for us is the case when there is some equivalence relation $\sim$ on A. Let

$$[a] = \{a' \in A \mid a \sim a'\}$$

denote the equivalence class of an element $a \in A$ and

$$A/\sim = \{[a] \mid a \in A\}$$

the set of all equivalence classes. Then, given two maps $f, g : A \to B$, we define the *quotient distance* as

$$\delta^{\sim}(f,g) = \frac{|\{\ [a] \in A/\sim \mid\ f|_{[a]} \neq g|_{[a]}\ \}|}{|A/\sim|}$$

Given a field $\mathbb{F}$, we write $\mathbb{F}[X]$ for the polynomial ring in an indeterminate variable X. More generally, we write $\mathbb{F}[X_1, \ldots, X_n]$ for the ring of polynomials in the multiple formal variables $X_1, \ldots, X_n$. We will always use capitalized Latin letters to denote the formal variables in polynomials. We denote $\mathbb{F}^{\leq d}[X]$ for the set of polynomials of degree at most d.

Given polynomials $f(X), g(X) \in \mathbb{F}[X]$, we write $f(X) = g(X)$ to indicate the formal equality of the polynomials—meaning that all of their coefficients in $\mathbb{F}$ are equal. In contrast, we write $f(x) = g(x)$ for the equality of the evaluations of the polynomials on a specific $x \in \mathbb{F}$. Formal equality of polynomials is a stricter notion than equality on evaluations in that $f(X) = g(X)$ always implies $f(x) = g(x)$ for all $x \in \mathbb{F}$. Note that in general the converse is false. For an example of this, consider the polynomials $f(X) = 0$ and $g(X) = X^2 - X$ in $\mathbb{F}_2[X]$. Then $F(X) \neq G(X)$ but $f(x) = g(x)$.

Let $\mathbb{F}$ be a field of characteristic p. A partial converse is true when we restrict to polynomials of degree strictly less than p. Let $f(X), g(X) \in \mathbb{F}^{\leq d}[X]$, where $d < p$, and suppose $f(x) = g(x)$ on at least $d + 1$ points; then $f(X) = g(X)$. We can rephrase this statement using the language of distance:

$$d(f,g) \leq 1 - \frac{d+1}{|\mathbb{F}|} \implies f(X) = g(X). \tag{4}$$

This follows from interpolation: $d + 1$ points uniquely specify a degree $\leq d$ polynomial. This property means that polynomials in $\mathbb{F}^{\leq d}[X]$ have a minimum distance between them, a fact which will play a crucial role in the sequel.

A special case of Relative Hamming Distance that is worth clarifying is the context where the codomain is a polynomial ring $\mathbb{F}[Y]$. This situation manifests in the context of bivariate polynomials $F \in \mathbb{F}[X, Y]$, where we evaluate in one of the variables, e.g. X, to produce a formal univariate polynomial in the other, Y. We denote this function as $F(-, Y) : \mathbb{F} \to \mathbb{F}[Y]$. Given two bivariate polynomials $F, G \in \mathbb{F}[X, Y]$, the Relative Hamming Distance between $F(-, Y)$ and $G(-, Y)$ is given as

$$d(F(-,Y), G(-,Y)) = \mathbb{P}_{x \in \mathbb{F}}[F(x,Y) \neq G(x,Y)],$$

where the non-equality $F(x, Y) \neq G(x, Y)$ is of formal polynomials in $\mathbb{F}[Y]$ whose coefficients are evaluations in x.

1.3 Coding Theory and Reed-Solomon Codes

We first briefly discuss the setup of coding theory, which is roughly the study of encoding and decoding data of one form into that of another. More precisely, consider a set A of inputs, a metric space (B, d) for the output *language*, a subset $B^* \subset B$ of *codewords*, and an (intended) *encoding* map $\chi : A \to B$ that is injective and has image B^*. The idea is that the actual process of encoding may be noisy or error-prone, and hence its outcome $b \in B$ might not be a codeword in B^*. The problem of *decoding* is deducing the likeliest $a \in A$ from which b was encoded. This is typically done by selecting, when it exists, the *unique decoding* $\hat{b}$ of b, defined as

$$\hat{b} = \arg\min_{b' \in B} d(b, b'),$$

and then decoding b as $\chi^{-1}(\hat{b}) \in A$. The set B^* is thus chosen to facilitate this process, while satisfying other constraints, such as efficiency. In particular, for the decoding process to work as hoped, it is desirable for any two distinct codewords to be sufficiently far apart. We define the *minimum distance* δ of B^* as the minimum distance between any two distinct codewords, which means that for any $b, b' \in B^*$

$$d(b, b') < \delta \implies b = b'.$$

The *unique decoding radius* of the language (B, d) is the quantity ϵ that guarantees that when $d(b, B^*) < \epsilon$, there is a unique decoding $\hat{b}$. It is always the case that $\epsilon \geq \frac{1}{2}\delta$, as, if there exists b and two codewords $\hat{b}, \hat{b}' \in B^*$ such that

$$d(b, \hat{b}) = d(b, \hat{b}') < \frac{1}{2}\delta$$

then the triangle inequality would imply the violation of minimum distance:

$$\begin{aligned} d(\hat{b}, \hat{b}') &\leq d(\hat{b}, b) + d(b, \hat{b}') \\ &< \delta \end{aligned}$$

We now specialize this setup to the context relevant in the paper. Let $\mathbb{F}$ be a field and $L \subset \mathbb{F}$. We define our language to be $\mathbb{F}$ evaluations on L, i.e. maps $f : L \to \mathbb{F}$, with distance given by Relative Hamming Distance. We now define the relevant set of codewords as follows.

Definition 1.1. *The* Reed-Solomon Code $RS[\mathbb{F}, L, d]$ *is defined as the set of evaluations on* L *of polynomials on* $\mathbb{F}$ *of degree at most* d*.*

$$RS[\mathbb{F}, L, d] = \{f|_L \mid f \in \mathbb{F}^{\leq d}[x]\}$$

The *rate* of $RS[\mathbb{F}, L, d]$ is defined as $\rho = d/|L|$. Note that $\rho \in [0, 1]$, since any codeword $f : L \to \mathbb{F}$ can be recovered, via interpolation, as a polynomial of degree less than $|L|$. Hence we typically choose d much smaller than $|L|$. We can compute the minimum distance by recalling (4):

$$d(f, g) > \frac{|L| - d}{|L|} = 1 - \rho,$$

which means that the unique decoding radius is bounded by $\frac{1}{2}(1 - \rho)$.

1.4 Folding Polynomials

Fix some finite field $\mathbb{F}$ and let $N = |\mathbb{F}| - 1$. We cite the following fact.

Theorem 1.1. *The multiplicative group $\mathbb{F}^\times = \mathbb{F} \setminus \{0\}$ is cyclic, i.e. $\mathbb{F}^\times \cong \mathbb{Z}/N\mathbb{Z}$.*

Let $s \in \mathbb{N}$ such that $s|N$. Since $1 \mapsto N/s$ is an embedding $\mathbb{Z}/s\mathbb{Z} \hookrightarrow \mathbb{Z}/N\mathbb{Z}$, this theorem implies the existence of an order s cyclic subgroup $\Gamma = \langle\gamma\rangle \hookrightarrow \mathbb{F}^\times$. Thus γ is an s^{th} primitive root of unity meaning $\gamma^s = 1$, but, for $l \in \underline{s}^*$, $\gamma^l \neq 1$. Hence for all $l \in \underline{s}^*$.

$$\sum_{j=0}^{s-1} \gamma^{lj} = \frac{\gamma^{ls} - 1}{\gamma^l - 1} = 0. \tag{5}$$

Given $k \in \underline{s}$, we can use γ to define a projection[1] $\pi_k^s : \mathbb{F}[X] \to \mathbb{F}[X^s]$ which extracts monomials of the form $a_j X^j$ for which $j \equiv k \mod s$ and then divides by X^k. Explicitly if $p(X) = \sum_{i=0}^n a_i X^i$

$$\pi_k^s(p(X)) = \sum_{i=0}^{\lfloor \frac{n-k}{s} \rfloor} a_{k+si} X^{si} = \frac{1}{s} \sum_{j=0}^{s-1} (\gamma^j X)^{-k} p(\gamma^j X)$$

This second equality is important as it means that $\pi_k(p(X))$ can be computed purely from evaluations of $p(X)$. To see why this equality holds, let's apply π_k^s to a monomial X^r.

$$\begin{aligned}
\pi_k^s(X^r) &= \frac{1}{s} \sum_{j=0}^{s-1} (\gamma^j X)^{-k} (\gamma^j X)^r \\
&= \frac{X^{r-k}}{s} \sum_{j=0}^{s-1} \gamma^{(r-k)j} \\
&= \begin{cases} X^{r-k} & r \equiv k \mod s \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

where the otherwise case is deduced from (5). Linearly extending this result gives the desired behavior for π_k^s. As $\pi_k^s(p(X)) \in \mathbb{F}[X^s]$, we can find a polynomial $p_k(X) \in \mathbb{F}[X]$ such that $p_k(X^s) = \pi_k^s(p(X))$. If $p(X)$ is of degree d, this means that $p_k(X)$ has degree at most $\lfloor d/s \rfloor$. From here, we can define, given some $\alpha \in \mathbb{F}$, the s-radix fold operation ∇_α^s as follows.

$$\nabla_\alpha^s(p)(X) = \sum_{k=0}^{s-1} \alpha^k p_k(X). \tag{6}$$

To study the behavior folding, it will be useful to introduce the *coset distance*, a special case of the quotient distance when the equivalence relation is given by quotienting by $\Gamma = \langle\gamma\rangle$:

$$\delta^\Gamma(f, g) = \frac{|\{\Gamma a \in L/\Gamma \mid g|_{\Gamma a} \neq h|_{\Gamma a}\}|}{|L/\Gamma|}$$

[1] Calling π_k^s a projection is technically an abuse of notation, since we divide by X^k.

and observe that

$$d(\nabla_\alpha^s f, \nabla_\alpha^s g) \le \delta^\Gamma(f, g).$$

Similar to a fold is a *roll*, where one simply combines unrelated polynomials $q_0, \ldots, q_{n-1}$ into a random sum

$$\nabla_\alpha([q_0, \ldots, q_{n-1}])(X) = \sum_{k=0}^{n-1} \alpha^k q_k(X) \tag{7}$$

Most often, we will have combined folds and rolls, where the fold of p will be combined with an array of unfolded $q_0, \ldots, q_{n-1}$ as follows.

$$\nabla_\alpha^s(p, [q_0, \ldots, q_{n-1}])(X) = \sum_{k=0}^{s-1} \alpha^k p_k(X) + \sum_{k=0}^{n-1} \alpha^{s+k} q_k(X) \tag{8}$$

When considering combined folds and rolls, coset distance is less useful and instead we talk about *correlated agreement* as in [1].

Definition 1.2 (Correlated Agreement)**.** *Let $\boldsymbol{f} = (f_0, \cdots, f_l)$ be a set of functions $L \to \mathbb{F}$ and $RS = RS[\mathbb{F}, L, d]$ a Reed-Solomon code. The correlated agreement, $C(\boldsymbol{f}, RS)$, is the maximal normalized size of a subset $L' \subset L$ such that $\boldsymbol{f}|_{L'} \subset RS' = RS[\mathbb{F}, L', d]$. Explicielty,*

$$C(\boldsymbol{f}, RS) = \frac{1}{|L|} \max_{\substack{L' \subset L \\ \boldsymbol{f}|_{L'} \in RS'}} |L'|$$

As explained in Lemma 3.2, if f is a codeword far from the RS code, the correlated agreement of projections of f will be small. Additionally, Theorem 3.3, proven in [1] shows that combining a collection of function with small correlated agreement will, with high probability, produce a function far from the RS code.

2 The Protocol

Given an index set $\underline{I}$, consider a set of codewords

$$f_I = \{f_i : L_i \to \mathbb{F}\}_{i \in \underline{I}},$$

in which each $f_i \in f_I$ is purported to be a Reed-Solomon codeword of degree d_i. Furthermore, suppose they satisfy the following criteria

- $d_i | d_{i-1}$ for $i \in I^*$
- $L_0 = \langle \omega \rangle$
- $L_i = \langle \omega^{d_0/d_i} \rangle$ for all $i \in I$

We will describe a variant of the FRI protocol, which we name *multi-FRI*. The idea of the protocol is to probabilistically check whether the codewords in f_I are all indeed Reed-Solomon of specified degree. The protocol has two phases: a *commit phase* and a *query phase*.

2.1 Commit Phase

The protocol begins with the prover $\mathbf{P}$ sending f_I to the verifier $\mathbf{V}$. The commit phase proceeds across $\underline{R}$ rounds, where at the beginning of round $r \in \underline{R}$, $\mathbf{P}$ begins with a codeword $\varphi^{(r)}$, $\mathbf{V}$ sends a random value $\alpha^{(r)} \in \mathbb{F}$ to $\mathbf{P}$, and then $\mathbf{P}$ sends a new codeword $\varphi^{(r+1)}$, which is purportedly some fold or roll $\nabla^{(r)}$, but, if $\mathbf{P}$ is malicious, could be any code.

More precisely, in the initialization step—a sort of round -1 perhaps—$\mathbf{V}$ sends $\alpha^{(-1)} \in \mathbb{F}$ and $\mathbf{P}$ is to initialize $\varphi^{(0)}$ by taking a pure roll of all codewords $f \in f_I$ for which $\deg f = d_0$

$$\varphi^{(0)} = \nabla_{\alpha^{(0)}}([f \in f_I \mid \deg f = f_0]),$$

where $\varphi^{(0)} = f_0$ in the case that no other $f \in f_I$ has degree d_0. By construction, $\deg \varphi^{(0)} = d_0$. This initialization is not itself really a round of the protocol, as it does not have any starting codeword and does not involve any decrementing of the codeword degree.

For each actual round $r \in \underline{R}$, $\mathbf{P}$ begins with $\varphi^{(r)} : L^{(r)} \to \mathbb{F}$ of purported degree $d^{(r)}$, $\mathbf{V}$ sends $\alpha^{(r)}$, and $\mathbf{P}$ is to send $\varphi^{(r+1)} : L^{(r+1)} \to \mathbb{F}$ via performing a fold or roll $\nabla^{(r)} = \nabla_{\alpha^{(r)}}^{\Gamma^{(r)}}$ of radix-$s^{(r)}$ using a degree $s^{(r)}$ cyclic group $\Gamma^{(r)}$, to produce a codeword of purported degree

$$d^{(r+1)} = \frac{d^{(r)}}{s^{(r)}}.$$

Letting $S^{(r)} = \prod_{j=0}^{r-1} s^{(j)}$, this means $\varphi^{(r)}$ will have domain $L^{(r)} = \langle \omega^{S^{(r)}} \rangle$ and degree $d^{(r)} = d_0/S^{(r)}$. Note that the rate $\rho = d^{(r)}/|L^{(r)}|$ remains constant.

For $i \in I^\star$, there is a round $r_i \in \underline{R}$ in which the fold also includes the roll all the f_i with degree d_i, given by the set $f_{\langle d_i \rangle} = [f \in f_I \mid \deg f = d_i]$ with $|f_{\langle d_i \rangle}| = n^{(r_i)}$, yielding

$$\varphi^{(r_i+1)} = \nabla_{\alpha^{(r_i)}}^{\Gamma^{(r_i)}}(\varphi^{(r_i)}, f_{\langle d_i \rangle}),$$

which we write simply as $\varphi^{(r_i+1)} = \nabla^{(r)}\varphi^{(r_i)}$. Observe that $d^{(r_i+1)} = d_i$ and hence $L^{(r_i+1)} = L_i$.

2.2 Query Phase

In this phase, there is no more interaction, and $\mathbf{V}$ performs *inconsistency checks* across all rounds on T random samples $\mu_0, \ldots, \mu_{T-1} \in L^{(0)}$. In particular, for each μ_t, let $\mu_t^{(r)} = \mu_t^{S^{(r)}}$. Define $\mathbf{catch}^{(r)}(\mu_t^{(r)})$ as the event where $\mathbf{V}$ catches a value inconsistent with folding

$$\varphi^{(r+1)}(\mu_t^{(r+1)}) \neq \nabla^{(r)}(\varphi^{(r)})(\mu_t^{(r)})$$

The protocol rejects, i.e. outputs $\mathbf{REJ}$, if $\mathbf{catch}^{(r)}(\mu_t^{(r)})$ occurs on any round r for any sample μ_t. Otherwise, it accepts, i.e. returns $\mathbf{ACC}$.

2.3 Pseudocode

We can formalize the protocol in pseudocode as follows.

```
# Commit Phase
P sends f_I
P initializes φ^(0)
for r ∈ R:
    V sends α^(r) ∈ F
    P sends φ^(r+1)
P sends φ^(R) as a formal polynomial

# Query Phase
for t ∈ T:
    V samples μ_t ∈ L^(0)
    for r ∈ R:
        if catch^(r)(μ_t^(r)):
            return REJ
return ACC
```

3 Soundness Analysis

The prover **P** is specified by functions $\{\varphi^{(r)} : L^{(r)} \to \mathbb{F}\}_{r \in \underline{R}}$, where $\varphi^{(r)}$ depends on $\alpha_0, \ldots, \alpha_{r-1}$. If all of **P**'s input codewords f_i are Reed-Solomon, and if **P** indeed performs every round consistently as specified by the protocol, then, by construction, the event accept (**ACC**) that all of **V**'s checks pass is guaranteed to occur. This means that the algorithm enjoys the *completeness* property.

In this section, we are interested in the other scenario; in particular, when **P** starts with codewords that are not all Reed-Solomon, and attempts to fool **V** by not following the protocol exactly, but doing so in a manner that attempts to go undetected. By the *soundness* of the protocol, we mean an upper bound for $\mathbb{P}[\mathbf{ACC}]$ in the context of such a malicious prover. There are broadly two sources of soundness error.

1. Distortion (**DIST**) can occur in the commit phase, and corresponds to the event that, in some round, a codeword that starts off far from Reed-Solomon ends up much closer.
2. Nondetection (**NDCT**), in the case of non-distortion, can occur during one of T samples of the query phase, and corresponds to the event that, by unlucky choice of sample, **V**'s checks do not detect the fact that **P** started with input codewords that were not Reed-Solomon.

We bound the soundness by applying conditional probability analysis as follows

$$\mathbb{P}[\mathbf{ACC}] = \mathbb{P}[\mathbf{ACC} \mid \mathbf{DIST}]\mathbb{P}[\mathbf{DIST}] + \mathbb{P}[\mathbf{ACC} \mid \neg\mathbf{DIST}]\mathbb{P}[\neg\mathbf{DIST}] \tag{9}$$

$$\leq \mathbb{P}[\mathbf{DIST}] + \mathbb{P}[\mathbf{NDCT}]^T \tag{10}$$

noting that $\mathbb{P}[\mathbf{ACC} \mid \neg\mathbf{DIST}] = \mathbb{P}[\mathbf{NDCT}]^T$ since we run T independent checks. Furthermore, in practice, $\mathbb{P}[\neg\mathbf{DIST}]$ is close to 1 and $\mathbb{P}[\mathbf{ACC} \mid \mathbf{DIST}]$ is much larger than $\mathbb{P}[\mathbf{DIST}]$ so this is a reasonable approximation.

The probabilities of the events **DIST** and **NDCT** can be traded off based on precisely how one defines distortion. In particular, this tradeoff is determined by the choice of a parameter called the *distortion boundary* δ^*, which we use to formally define distortion. Let δ be the coset distance of the current codeword to RS at the start of a FRI round. Distortion occurs in this round if one of the following two events occur in the round's fold or roll.

1. if $\delta \geq \delta^*$, the relative hamming distance falls below δ^*.
2. if $\delta < \delta^*$, the relative hamming distance falls below δ.

In 3.1, we will show that, no matter the choice of δ^*, the following bound holds

$$\mathbb{P}[\mathbf{NDCT}] \leq 1 - \delta^*$$

In 3.2, we choose $\delta^* = \frac{1}{4}(1-\rho)^2$, and prove the bound

$$\mathbb{P}[\mathbf{DIST}] \leq \frac{|L|}{|\mathbb{F}|}$$

and thus obtain the soundness error

$$\mathbb{P}[\mathbf{ACC}] \leq \frac{|L|}{|\mathbb{F}|} + (1-\delta^*)^T. \tag{11}$$

For large fields where $\frac{|L|}{|\mathbb{F}|}$ is far smaller than the target soundness, it is beneficial to increase δ^* which will allow us to be more efficient by decreasing our number T of needed queries. In 3.3, we discuss, via citing Theorem 3.3 from [1], how the more complicated choice of distortion boundary $\delta^* = 1 - \sqrt{\rho}(1 + \frac{1}{2m})$ allows us to reach the current state of the art.

3.1 Nondetection

Given a single random sample $a^{(0)} \in L_0$, our goal is to compute the probability of nondetection (**NDCT**); though it will be easier to instead work with the probability of the complementary event **DET** of detection:

$$\mathbb{P}[\mathbf{NDCT}] = 1 - \mathbb{P}[\mathbf{DET}].$$

In this section, we prove the following theorem that will bound this probability.

Theorem 3.1. *Let δ^* be the distortion boundary. Then*

$$\mathbb{P}[\mathbf{DET}] \geq \min(\delta^*, \max_{i \in \underline{I}} d(f_i, RS_i))$$

In the case that $\max_{i \in \underline{I}} d(f_i, RS_i) > \delta^*$, this implies our desired bound.

$$\mathbb{P}[\mathbf{NDCT}] = 1 - \mathbb{P}[\mathbf{DET}] \leq 1 - \delta^*$$

Note that **DET** is precisely the event in which there exists a round where **V** catches that **P** submitted an inconsistent value

$$\mathbb{P}[\mathbf{DET}] = \mathbb{P}_{a \in L^{(0)}}[\exists r \in \underline{R} \mid \mathbf{catch}^{(r)}(a^{(r)})]$$

We argue that $\mathbf{catch}^{(r)}(a^{(r)})$ occurs in at most one round r. Otherwise, there are rounds $i < j$ in which the attacker fudged both $\varphi^{(i)}(a^{(i)})$ and $\varphi^{(j)}(a^{(j)})$. Hence we may just consider an equally effective attacker who only fudged $\varphi^{(j)}(a^{(j)})$. As we check chains of values $\varphi^{(0)}(a^{(0)}), \ldots, \varphi^{(R)}(a^{(R)})$, the probability of catching these attackers will be identical so it suffices to simply consider this second attacker. Therefore, by the disjointness of catching an inconsistency across rounds, and by the definition of Relative Hamming Distance

$$\begin{aligned}\mathbb{P}[\mathbf{DET}] &= \sum_{r=0}^{R-1} \mathbb{P}_{a^{(r)} \in L^{(r)}}[\mathbf{catch}^{(r)}(a^{(r)})] \\ &= \sum_{r=0}^{R-1} d(\nabla\varphi^{(r)}, \varphi^{(r+1)})\end{aligned}$$

where, for the sake of notational convenience, we denote by ∇ the roll and or fold at a given round, without adorning it with further symbol. We now prove a result that will do most of the work in establishing the above theorem.

Lemma 3.1. *Suppose for all rounds $r > j$, that $\varphi^{(r)}$ is within the distortion boundary. Then, assuming that distortion does not occur, we have*

$$\mathbb{P}[\mathbf{DET}] \geq d(\nabla\varphi^{(j)}, RS^{(j)})$$

Proof. As there is no distortion and $\varphi^{(j+1)}$ is within the distortion boundary

$$\begin{aligned}d(\varphi^{(j+1)}, RS^{(j+1)}) &\leq d(\nabla\varphi^{(j+1)}, RS^{(j+2)}) \\ &\leq d(\nabla\varphi^{(j+1)}, \varphi^{(j+2)}) + d(\varphi^{(j+2)}, RS^{(j+2)})\end{aligned}$$

By inducting up to round R, we have that

$$d(\varphi^{(j+1)}, RS^{(j+1)}) \leq \sum_{i=j+1}^{R-1} d(\nabla\varphi^{(i)}, \varphi^{(i+1)}) + d(\varphi^{(R)}, RS^{(R)}) \tag{12}$$

$$= \sum_{i=j+1}^{R-1} d(\nabla\varphi^{(i)}, \varphi^{(i+1)}) \tag{13}$$

where, by construction of the protocol, $d(\varphi^{(R)}, RS^{(R)}) = 0$. Thus,

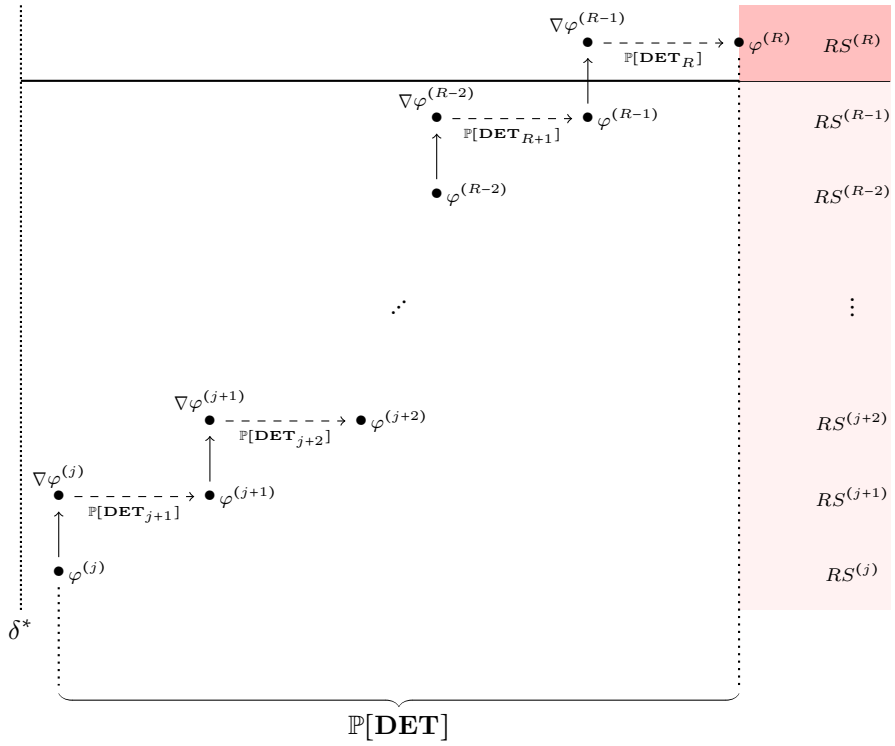
$$d(\nabla\varphi^{(j)}, RS^{(j+1)}) \leq d(\nabla\varphi^{(j)}, \varphi^{(j+1)}) + d(\varphi^{(j+1)}, RS^{(j+1)}) \tag{14}$$

$$\leq d(\nabla\varphi^{(j)}, \varphi^{(j+1)}) + \sum_{i=j+1}^{R-1} d(\nabla\varphi^{(i)}, \varphi^{(i+1)}) \tag{15}$$

$$= \sum_{i=j}^{R-1} d(\nabla\varphi^{(i)}, \varphi^{(i+1)}) \tag{16}$$

$$\leq \mathbb{P}[\mathbf{DET}] \qquad \square$$

Because a picture is worth a thousand words, we offer a visual rendering of the above proof, where we depict the best case scenario for the prover.



We split the nondetection analysis into two cases, depending on whether the distance from the Reed-Solomon family is ever outside of the distortion boundary δ^*.

1. Suppose that for all $r \in \underline{R}$, $\delta^{(r)} \leq \delta^*$. Let $\delta = \max_{i \in \underline{I}} d(f_i, RS_i)$ and f_i be the codewordthat realizes this maximum. Recall that f_i is rolled in round r_i. Noting that folds are non-decreasing within the distortion boundary,

and then citing Lemma 3.1 yields

$$\begin{aligned}\delta &= d(f_i, RS_i)\\ &\le d(\nabla\varphi^{(r_i)}, RS^{(r_i+1)})\\ &\le \mathbb{P}[\mathbf{DET}]\end{aligned}$$

2. Let j be the largest index such that $\delta^{(j)} > \delta^*$. Then by Lemma 3.1

$$\begin{aligned}\delta^* &\le d(\nabla\varphi^{(j)}, RS^{(j+1)})\\ &\le \mathbb{P}[\mathbf{DET}]\end{aligned}$$

This completes the proof of Theorem 3.1

3.2 Distortion

In this section we will prove Theorem 3.2, which states that in the case of pure folding with distortion boundary $\delta^* = \frac{1}{4}(1-\rho)^2$, that the probability of Distortion occurring on a given round, in which a radix-s pure fold occurs, can be bounded as follows

$$\mathbb{P}[\mathbf{DIST}] \le \frac{s-1}{s}\frac{|L|}{|\mathbb{F}|}$$

Therefore, given a multi-FRI protocol with radices $[s^{(r)}]_{r\in\underline{R}}$, we can bound the probability of distortion across the protocol by

$$\mathbb{P}[\mathbf{DIST}] \le \sum_{r=0}^{R-1} \frac{s^{(r)}-1}{s^{(r)}}\frac{|L^{(r)}|}{|\mathbb{F}|} \tag{17}$$

$$= \frac{|L|}{|\mathbb{F}|}\sum_{r=0}^{R-1}\frac{s^{(r)}-1}{S^{(r+1)}} \tag{18}$$

$$\le \frac{|L|}{|\mathbb{F}|}, \tag{19}$$

where the final step can be worked out via bounding by the sum of the infinite "geometric-like" series.

Although we only consider folding in this section, these methods will extend to the rolling case but require slightly more sophisticated math. This is completely dealt with in Section 4.1 of [1], whose bounds we discuss in Section 3.3.

Theorem 3.2. *Fix $f : L \to \mathbb{F}$ and order s cyclic subgroup $\Gamma \subset L$. Given some κ, we can define the drop set*

$$\mathcal{D}^s_\kappa(f) = \{\alpha \in \mathbb{F} \mid d(\nabla^s_\alpha f, RS^s) < \kappa\}$$

Let $\delta = \delta^\Gamma(f, RS)$. Then

(i) when $\delta < \delta^*$, distortion corresponds to the distance non-decreasing

$$\mathbb{P}[\mathcal{D}^s_\delta(f)] \leq \frac{(s-1)}{s}\frac{|L|}{|\mathbb{F}|}$$

(ii) when $\delta \geq \delta^*$, distortion corresponds to the distance not falling below δ^*

$$\mathbb{P}[\mathcal{D}^s_{\delta^*}(f)] \leq \frac{(s-1)}{s}\frac{|L|}{|\mathbb{F}|}$$

Proof.

(i) In this range, there is a unique Reed-Solomon decoding $\hat{f}$. For α to be in the drop set, there must exist some $a \in L^s$ for which $f|_{a\Gamma} \neq \hat{f}|_{a\Gamma}$, but $\nabla^s_\alpha f(a) = \nabla^s_\alpha \hat{f}(a)$. Since such α are points of intersection between two degree $s-1$ polynomials, there are at most $s-1$ such α. Taking the union bound across $a \in L^s$ gives at most $\frac{s-1}{s}|L|$ such α across all $\mathbb{F}$.

(ii) We demonstrate the claim via proving the contrapositive of a stricter version[2], where we replace $\delta \geq \delta^*$ with $\delta \geq \frac{1}{2}(1-\rho)$

$$\mathbb{P}[\mathcal{D}^s_{\delta^*}(f)] > \frac{(s-1)}{s}\frac{|L|}{|\mathbb{F}|} \implies \delta < \frac{1}{2}(1-\rho).$$

We will suppose the left hand bound, and use it to construct a unique decoding $\hat{f}$, which satisfies $\delta(f,\hat{f}) < \frac{1}{2}(1-\rho)$.

We do so via lifting our problem to bivariate polynomials in $\mathbb{F}[X,Y]$. In particular, we will use Theorem A.1 to construct such a polynomial $K(X,Y)$ with low degree in both variables, and use it to define the low degree single variable polynomial $\hat{f}(X) = K(X^s, X)$, which we show is close to f. We proceed as follows.

Let $F(X,Y)$ be defined on $(x,\alpha) \in L' \times \mathcal{D}^s_{\delta^*}(f)$ by

$$F(x,\alpha) = \nabla^s_\alpha f(x),$$

noting that for $x \in L$, $f(x) = F(x^s, x)$. By construction, F has degree

$$\deg_X F \leq |L'| - 1 \qquad \deg_Y F \leq s-1$$

Next, observe that, because $\delta^* < \frac{1}{2}(1-\rho)$, we have that $\alpha \in \mathcal{D}^s_{\delta^*}(f)$ implies the existence of a unique decoding $\widehat{\nabla^s_\alpha f}$ of $\nabla^s_\alpha f$ for all $\alpha \in \mathcal{D}^s_{\delta^*}(f)$. We can interpolate the $|\mathcal{D}^s_{\delta^*}(f)|$ points $(\alpha, (\widehat{\nabla^s_\alpha f})(x))$ to define $G(X,Y)$ on $(x,\alpha) \in L' \times \mathcal{D}^s_{\delta^*}(f)$ by

$$G(x,\alpha) = \left(\widehat{\nabla^s_\alpha f}\right)(x),$$

[2]This is an indication of the fact that with more care, these bounds can be improved.

which has degree

$$\deg_X G \le \frac{d}{s} - 1 \qquad \deg_Y G \le |\mathcal{D}^s_{\delta^*}(f)| - 1.$$

We now show that the relative Hamming distance $d(F, G)$ between F and G on the domain $L' \times \mathcal{D}^s_{\delta^*}(f)$ is bounded by δ^*:

$$\begin{aligned}
d(F,G) &= \mathbb{P}_{(x,\alpha)}[F(x,\alpha) \neq G(x,\alpha)] && (20)\\
&= \mathbb{P}_{(x,\alpha)}\left[\left(\nabla^s_\alpha f\right)(x) \neq \left(\widehat{\nabla^s_\alpha f}\right)(x)\right] && (21)\\
&= \frac{1}{|\mathcal{D}^s_{\delta^*}(f)|} \sum_{\alpha \in \mathcal{D}^s_{\delta^*}(f)} \mathbb{P}_x\left[\left(\nabla^s_\alpha f\right)(x) \neq \left(\widehat{\nabla^s_\alpha f}\right)(x)\right] && (22)\\
&= \frac{1}{|\mathcal{D}^s_{\delta^*}(f)|} \sum_{\alpha \in \mathcal{D}^s_{\delta^*}(f)} d(\nabla^s_\alpha f, \widehat{\nabla^s_\alpha f}) && (23)\\
&< \frac{1}{|\mathcal{D}^s_{\delta^*}(f)|} \sum_{\alpha \in \mathcal{D}^s_{\delta^*}(f)} \delta^* && (24)\\
&= \delta^* && (25)
\end{aligned}$$

Choose σ^*, τ^* such that $\delta^* = \sigma^*\tau^*$ and

$$\begin{aligned}
1 &> \sigma^* + \tau^* + \frac{d/s - 1}{|L'|} + \frac{s-1}{|\mathcal{D}^s_{\delta^*}(f)|} && (26)\\
&= \sigma^* + \tau^* + \rho && (27)
\end{aligned}$$

Where we use the facts that $|L'| = \frac{|L|}{s}$ and our initial assumption on $\mathbb{P}[\mathcal{D}^s_{\delta^*}(f)]$ which is equivalent to

$$\frac{s-1}{|\mathcal{D}^s_{\delta^*}(f)|} < \frac{s}{|L|}.$$

Applying Theorem A.1 to $F(X,Y), G(X,Y)$ yields a $K(X,Y)$ with degree

$$\deg_X K \le \frac{d}{s} - 1 \qquad \deg_Y K \le s - 1$$

satisfying

$$d(K(-,Y), F(-,Y)) \le \sigma^*.$$

Now define $\hat{f}(X) = K(X^s, X)$, which has low degree:

$$\begin{aligned}
\deg \hat{f} &= s(\deg_X K) + \deg_Y K\\
&\le s\left(\frac{d}{s} - 1\right) + s - 1\\
&= d - 1.
\end{aligned}$$

It remains to show that $\delta(f, \hat{f}) < \frac{1}{2}(1-\rho)$. Note that when $x \in L$ is such that $F(x^s, Y) = K(x^s, Y)$ as polynomials in $\mathbb{F}[Y]$, this implies that

$$f(x) = F(x^s, x) = K(x^s, x) = \hat{f}(x).$$

We can thus reason as follows

$$\begin{aligned}
\delta(f, \hat{f}) &= 1 - \mathbb{P}_x[f(x) = \hat{f}(x)] && (28)\\
&\leq 1 - \mathbb{P}_x[F(x, Y) = K(x, Y)] && (29)\\
&\leq \sigma^* && (30)\\
&< \frac{1}{2}(1-\rho). && (31)
\end{aligned}$$

Therefore $\hat{f}$ is the unique decoding of f.

□

3.3 More Optimal Bounds

In practice, the type of attack which we would like to prevent is one where an attacker attempts to prove an incorrect claim. In such cases, the attacker will start with a collection of codewords $\{f_i : L_i \to \mathbb{F}\}_{i \in \underline{I}}$ at least some of which are far from the corresponding Reed-Solomon Code $RS[\mathbb{F}, L_i, d_i]$. Mathematically, this means that we can assume

$$\max_{i \in \underline{I}} d(f_i, RS_i) \approx 1 - \rho + O\left(\frac{1}{|\mathbb{F}|}\right).$$

Now recall Definition 1.2, for correlated agreement. Observe that for any subset $\mathbf{g} \subset \mathbf{f}$ there is the trivial bound

$$C(\mathbf{f}, RS) \leq C(\mathbf{g}, RS) \tag{32}$$

and for singleton sets

$$C(\{f_0\}, RS) = 1 - d(f_0, RS). \tag{33}$$

The reason for discussing correlated agreement is that in general, given a codeword f with $d(f, RS)$ large, the projections $\pi_k^s(f)$ might be individually close to RS' but the correlated agreement for the set of projections must be small, as shown by Lemma 3.2. Hence instead of dealing with both a roll and a fold, we consider simply rolling together new functions with our projections. Then [1] shows that, given a set of functions with small enough correlated agreement, with high probability the roll of those functions will be far from an RS code.

Lemma 3.2. *Let $f : D \to \mathbb{F}$ be a function with $d(f, RS) > 1 - \alpha$. Fix an integer s dividing $|D|$ and let $\boldsymbol{u}$ denote the set of projections $\pi_k^s(f) : D^s \to \mathbb{F}$ for $k \in \underline{s}$. Then*

$$C(\boldsymbol{u}, RS') < \alpha$$

Proof. We show this via contradiction and so assume that $C(\mathbf{u}, RS') > \alpha$. This means we can find a subset $E \subset D^s$ with $\frac{|E|}{|D^s|} > \alpha$ and a collection of codewords $v_k \in RS'$ such that

$$\pi_k^s(f)|_E = v_k|_E.$$

Now consider the polynomial

$$v(x) = \sum_{i=0}^{s-1} x^i v_i(x^s) \in RS.$$

By construction, v and f agree on any point x such that $x^s \in E$. But this is a set of size $s|E|$ and so

$$d(f, RS) < d(f, v) < 1 - \frac{s|E|}{|D|} < 1 - \alpha$$

yielding the desired contradiction. □

We now state without proof the following theorem from [1].

Theorem 3.3. *Fix*

$$\delta^* = \sqrt{\rho}\left(1 + \frac{1}{2m}\right)$$

where m is chosen such that $m^7 > |L_0|^2$. Let $\boldsymbol{f} = (f_0, \cdots, f_l)$ be a set of functions, all of type $L^{(i)} \to \mathbb{F}$ for a fixed i. Then

$$\mathbb{P}_{\alpha \in \mathbb{F}}\left[d(\nabla_\alpha(\boldsymbol{f}), RS^{(i)}) < 1 - \max(C(\boldsymbol{f}, RS^{(i)}), \delta^*)\right] < \frac{(m+\frac{1}{2})^7 |L^{(i)}|^2}{3\rho^{\frac{3}{2}}|\mathbb{F}|} l.$$

Combining the Theorem 3.3 with Lemma 3.2 and equation (32), (33) shows that with probability at least[3]

$$\frac{(m+\frac{1}{2})^7 |L^{(i)}|^2}{3\rho^{\frac{3}{2}}|\mathbb{F}|}(s^{(i)} + n^{(i)} - 1)$$

$$d(\varphi^{(i)}, RS^i) \leq \min(\delta^*, \max_{f \in f_{\langle d_i \rangle}}(d(f, RS^i), d(\varphi^{(i-1)}, RS^{i-1}))) \tag{34}$$

Thus, letting i denote the final index where there exists an $f \in f_{\langle d_i \rangle}$ with $d(f, RS^i) > \delta^*$, we can bound the distortion risk by

$$\mathbb{P}[\mathbf{DIST}] = \frac{(m+\frac{1}{2})^7}{3\rho^{\frac{3}{2}}|\mathbb{F}|} \sum_{j=i}^{R} |L^{(j)}|^2 (s^{(j)} + n^{(j)} - 1) \tag{35}$$

$$\leq \frac{(m+\frac{1}{2})^7 |L^{(0)}|^2}{3\rho^{\frac{3}{2}}|\mathbb{F}|}\left(\sum_{j=0}^{R} \frac{s^{(j)} - 1}{\prod_{k=0}^{j-1}(s^{(k)})^2} + \sum_{j \in \underline{I}^*} \frac{|L_j|^2}{|L^{(0)}|^2}\right) \tag{36}$$

$$\leq \frac{(m+\frac{1}{2})^7 |L^{(0)}|^2}{3\rho^{\frac{3}{2}}|\mathbb{F}|}\left(\max_j s^{(j)} + \sum_{j \in \underline{I}^*} \frac{|L_j|^2}{|L^{(0)}|^2}\right) \tag{37}$$

[3]Note that $n^{(i)} = 0$ unless, for some j, $i = r_j$, where we roll in new functions on round i

This last step is a loose bound and could be improved if more information about the $s^{(j)}$ is known. If all $s^{(j)}$ are equal to s then it is $(s-1)+\frac{1}{s+1}$ which for $s=2$ recovers the distortion bound in [1]. Similarly, if when rolling, we choose an extra bit of randomness for each function we roll in, this extra factor $\sum_{j\in \underline{I}^*} \frac{|L_j|^2}{|L^{(0)}|^2}$ can be removed. Otherwise it can always be bounded by $|I|^*$ but this is a weak bound, particularly if much of the rolling happens after a few folds.

Hence as our distortion boundary is $\delta^* = 1-\sqrt{\rho}(1+\frac{1}{2m})$ we conclude that a better soundness bound for the multi-FRI protocol is:

$$\frac{(m+\frac{1}{2})^7|L^{(0)}|^2}{3\rho^{\frac{3}{2}}|\mathbb{F}|}\left(\max_j s^{(j)} + \sum_{j\in \underline{I}^*}\frac{|L_j|^2}{|L^{(0)}|^2}\right) + \left(\sqrt{\rho}\left(1+\frac{1}{2m}\right)\right)^T$$

Note that there is a natural trade-off here in how we choose m and in general it should be picked so that the two terms have similar size.

A Multivariate Polynomials

While much of the intuition developed from studying single variable polynomials will carry over into the multi-variable case, there are some important differences and some new results which emerge. This is a relatively brief overview, further information can be found in [2] or [1].

We start with a simple but fundamental result in the field, namely how many evaluations it takes to determine a multivariate polynomial.

Lemma A.1 (Multivariate Interpolation)**.** *Given 2 subsets $\mathbf{X} = x_0 \cdots x_d, \mathbf{Y} = y_0 \cdots y_e$ and, for each pair $(x_i, y_j) \in \mathbf{X}\times\mathbf{Y}$, a value $z_{i,j} \in \mathbb{F}$. There is a unique 2 variable polynomial $F(X,Y) \subset \mathbb{F}[X,Y]$ of degree $\leq (d,e)$ such that $F(x_i,y_j) = z_{i,j}$.*

Proof. This is a natural generalisation of the usual method of polynomial interpolation in one variable. For each fixed $y_j \in \mathbf{Y}$, univariate interpolation over $\mathbf{X}$ gives a unique polynomial $p_j(X) \in \mathbb{F}[X]$ of degree $\leq d$ such that $p_j(x_i) = z_{i,j}$. Let

$$p_j(X) = \sum_{k=0}^{d} a_{j,k}X^k$$

then, for each fixed k, by interpolating over $\mathbf{X}$ we can find a unique polynomial $q_k(Y) \in \mathbb{F}[Y]$ of degree $\leq e$ such that $q_k(y_j) = a_{j,k}$. Then

$$F(X,Y) = \sum_{k=0}^{d} q_k(Y)X^k$$

is the polynomial we are looking for. □

Observe that this naturally extends to interpolation for polynomials in k variables. The key takeaway is that while in 1 dimension any $d+1$ points

determines a d dimensional polynomial, in k dimensions, a $(d_1+1,\cdots,d_k+1)$ dimensional grid determines a $(d_1,\cdots,d_k)$ degree polynomial.

Moving on, one remarkable difference in the multivariate space is that there is far more freedom in vanishing polynomials.

Lemma A.2 (Lemma 4.2.13, [2]). *Let $S \subset \mathbb{F}\times\mathbb{F}$ have order $|S| \le ab$. Then there exists a nonzero polynomial $E(X,Y) \in \mathbb{F}[X,Y]$ of degree $\le (a,b)$ which vanishes on all of S.*

Proof. Let T be the space of all polynomials with X degree less than or equal to a and Y degree less than or equal to b. Consider the evaluation map $\phi : T \to \mathbb{F}^{|S|}$ which sends a polynomial $E(X,Y)$ to the vector

$$\phi(E(X,Y)) = [E(x_1,y_1),...,E(x_{|S|},y_{|S|})]$$

where $[(x_1,y_1),...,(x_{|S|},y_{|S|})]$ are an enumeration of the points in S. As the dimension of T is $(a+1)(b+1) > |S|$, this map has nontrivial kernel and so there exists a nonzero polynomial which maps to 0. □

Finally we have the following useful lemma which grantees that if a function $f(X,Y)$ on $\mathbf{X}\times\mathbf{Y}$ has low degree on all slices $x\times\mathbf{Y}$ and $\mathbf{X}\times y$, then it is itself low degree.

Proposition A.1 (Proposition 4.2.9, [2]). *Assume that $f(X,Y)$ is a function on $\mathbf{X}\times\mathbf{Y}$ such that for all $y\in\mathbf{Y}$, $f(X,y)$ agrees with some degree d polynomial in X and for all $x\in\mathbf{X}$, $f(x,Y)$ agrees with a degree e polynomial in Y.*

Then $f(X,Y)$ agrees with a polynomial $P(X,Y)$ of degree (d,e).

Proof. Pick an arbitrary set $\{y_0,...,y_e\}$ of $e+1$ points in $\mathbf{Y}$, and for each $j \in \underline{e+1}$, let δ_j be the e degree polynomial, satisfying

$$\delta_j(y_k) = \begin{cases} 1 & j = k \\ 0 & \text{else.} \end{cases}$$

Additionally, by assumption, for each y_k there is a degree d polynomial $p_k(x)$ with $f(X,y_k) = p_k(X)$ on $\mathbf{X}$.

Hence define $P(X,Y) = \sum_{i=0}^{e} \delta_i(Y)p_i(X)$. It remains to show that $P(X,Y) = f(X,Y)$ on $\mathbf{X}\times\mathbf{Y}$. For any $x\in\mathbf{X}$, we automatically know that $P(x,y_k) = f(x,y_k)$ for all $k \in \underline{e+1}$. Hence $P(x,Y)$ and $f(x,Y)$ agree on $e+1$ points and as both are polynomials of degree at most e this proves they are equal. □

A.1 Two Variable Divisibility Criterion

The main goal of this subsection will be to prove Lemma A.4. The overarching idea is that, given 2 polynomials $E(X,Y), P(X,Y) \in \mathbb{F}[X,Y]$, if we can find enough $x,y\in\mathbb{F}$ such that $E(x,Y)|P(x,Y)$ in $\mathbb{F}[Y]$ and $E(X,y)|P(X,y)$ in $\mathbb{F}[X]$, then $E(X,Y)|P(X,Y)$ in $\mathbb{F}[X,Y]$. To prove this we will need to introduce a more powerful algebraic gadget known as the resultant.

Definition A.1 (The Resultant)**.** *Given two polynomials f, g over an integral domain, we can write*

$$f(X) = \sum_{i=0}^{r} f_i X^i \qquad g(X) = \sum_{i=0}^{s} g_i X^i$$

then form the $r+s$ square matrix

$$M(f,g) = \begin{pmatrix} f_r & f_{r-1} & \cdots & \cdots & f_0 & 0 & \cdots & 0 \\ 0 & f_r & \cdots & \cdots & f_1 & f_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & f_r & \cdots & \cdots & f_1 & f_0 \\ g_s & \cdots & g_1 & g_0 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & & & & \ddots & & \vdots \\ \vdots & & \ddots & & & & \ddots & \vdots \\ 0 & \cdots & & 0 & g_s & \cdots & g & g_0 \end{pmatrix}$$

The resultant $R(f,g)$ it the determinant of $M(f,g)$.

The key feature that we will be using is that the resultant can determine if f and g share a common factor.

Lemma A.3. *The resultant $R(f,g) = 0$ if and only if f and g have a common factor. Indeed, passing to an algebraic closure and denoting the roots as λ_i and μ_j respectively (allowing for multiplicity),*

$$R(f,g) = f_r^s g_s^r \prod_{\substack{1 \le i \le r \\ 1 \le j \le s}} (\lambda_i - \mu_j) = f_r^s \prod_{1 \le i \le r} g(\lambda_i) = (-1)^{nm} g_s^r \prod_{1 \le j \le s} f(\mu_j).$$

The final equation of this lemma is often given as the definition of the resultant. We will use the resultant by considering $f, g \in \mathbb{F}[X,Y]$ as polynomials in Y (resp X) over the integral domain $\mathbb{F}[X]$ (resp $\mathbb{F}[Y]$). Then by counting 0's with multiplicity we will conclude that the resultant must be the 0 polynomial. To find the multiplicity of a 0, we will need the following simple proposition about the interaction of the determinant operation with derivatives.

Proposition A.2. *Given a collection of functions $f_{i,j} : \mathbb{F} \to \mathbb{F}$ let R denote the determinant $|M|$ with*

$$M = \begin{pmatrix} f_{1,1} & \cdots & f_{1,n} \\ \vdots & \ddots & \vdots \\ f_{n,1} & \cdots & f_{n,n} \end{pmatrix}.$$

Then

$$R^{(1)} = R' = \begin{vmatrix} f'_{1,1} & \cdots & f'_{1,n} \\ \vdots & \ddots & \vdots \\ f_{n,1} & \cdots & f_{n,n} \end{vmatrix} + \cdots + \begin{vmatrix} f_{1,1} & \cdots & f_{1,n} \\ \vdots & \ddots & \vdots \\ f'_{n,1} & \cdots & f'_{n,n} \end{vmatrix}$$

and more generally $R^{(n)}$ consists of a sum of determinants with at most n rows changed from M.

Explicit formula for $R^{(n)}$ can be constructed via the use of Multinomials. Let us now turn to the main lemma we want to show.

Lemma A.4 (Lemma 4.3, [1]). *Let $E(X,Y) \in \mathbb{F}[X,Y]$ be a polynomial with degree (E_X, E_Y) and $H(X,Y) \in \mathbb{F}[X,Y]$ have degree $\leq (E_X + K_X, E_Y + K_Y)$, where $K_X, K_Y \geq 0$. Assume that for m_X distinct points $x_1, \cdots, x_{m_X}$ and m_Y distinct points $y_1, \cdots, y_{m_Y}$*

$$\begin{aligned} H(x_i, Y) &= Q_{i,Y}(Y)E(x_i, Y) \qquad \deg_Y Q_{i,Y} \leq K_X \\ H(X, y_j) &= Q_{X,j}(X)E(X, y_j) \qquad \deg_Y Q_{X,j} \leq K_Y \end{aligned}$$

Assuming furthermore that

$$1 > \frac{E_X + K_X}{m_X} + \frac{E_Y + K_Y}{m_Y} \tag{38}$$

we show that $E(X,Y)$ divides $H(X,Y)$ as polynomials in $\mathbb{F}[X,Y]$. Additionally, letting $K(X,Y) = \frac{H(X,Y)}{E(X,Y)}$, K has degree $\leq (K_X, K_Y)$ and whenever $E(x_i, Y)$ (resp $E(X, y_j)$) is nonzero, $K(x_i, Y) = Q_{i,Y}(Y)$ (resp $K(X, y_j) = Q_{X,j}(X)$).

Proof. The idea of the proof will be by descent. We will show that if E and H satisfy the relevant inequalities then they share a common factor F. Assuming there is a common factor, we show that $\frac{E}{F}$ and $\frac{H}{F}$ also satisfy the requirements of the lemma and so by descent we can continue to remove common factors until $E = 1$. This would prove that E divides F.

We will actually do these steps in reverse. E.g. we start by showing that if E and P share a common factor, then the lemma will still apply to $\frac{E}{F}$ and $\frac{H}{F}$ with the points x_i, y_j where $F(x_i, Y) = 0$ or $F(X, y_j) = 0$ removed.

Hence assume that F has degree (F_X, F_Y). If $E = F$ then we are done. Similarly, if $F_X = E_X$ or $F_Y = E_Y$, the proof collapses to the one dimensional case which is easy. Hence assume that $(F_X, F_Y) < (E_X, E_y)$. Then looking just at degree, dividing by F will subtract F_X from the X degree and we will have to remove at most F_X points. As $\frac{E_X + K_X}{m_X} < 1$ we find that

$$\frac{E_X + K_X - F_X}{m'_X} < \frac{E_X + K_X - F_X}{m_X - F_X} < \frac{E_X + K_X}{m_X}.$$

Identical analysis works for Y and so we see that after dividing through we will still satisfy the same inequality. Hence assume that E and H share no common factors.

Now consider E, P as polynomials in Y over the integral domain $\mathbb{F}[X]$. If P has Y degree $< E_Y + H_Y$, formally add $0\ Y^{E_Y + H_Y}$ and henceforth treat it as a full $E_Y + H_Y$ degree polynomial. Then take the resultant $R(E,P)(X)$, this is a polynomial in X and directly from the definition we can see that the X degree is $\leq E_X(E_Y + H_Y) + E_Y(E_X + H_X) = D$. Note that D is symmetric in X, Y so we could have equivalently looked at the resultant as a function of Y. We will show that $R(E,P)$ has too many zeroes which implies it must be the 0 polynomial.

Pick any x_i r45t65f values. Then as $H(x_i, Y) = Q_{i,Y}(Y)E(x_i, Y)$ with $\deg_Y Q_{i,Y} \leq K_X$, all E_Y the rows in $M(E,P)(x_i)$ corresponding to $H(x_i, Y)$ coefficients are a linear combination of at most K_X+1 of the rows corresponding to $E(x_i, Y)$ coefficients. Hence the overall rank of $M(E,P)(x_j)$ is $\leq E_Y + P_Y$ and so R has a 0 of order E_Y at x_i. Hence this polynomial has at least $E_Y m_X$ 0's and so we must have $E_Y m_X \leq D$. Repeating the same argument switching the roles of X and Y we similarly conclude that $E_X m_Y \leq D$. Thus

$$D = E_X(E_Y + H_Y) + E_Y(E_X + H_X) \tag{39}$$

$$= \frac{(E_Y + H_Y)}{m_Y} E_X m_Y + \frac{(E_X + H_X)}{m_X} E_Y m_X \tag{40}$$

$$\leq \frac{(E_Y + H_Y)}{m_Y} D + \frac{(E_X + H_X)}{m_X} D \tag{41}$$

$$= \left(\frac{(E_Y + H_Y)}{m_Y} + \frac{(E_X + H_X)}{m_X} \right) D \tag{42}$$

but from our initial inequality we know the right hand side is strictly less than D unless $D = 0$. Hence $D = 0$ and so we have found a root and so we conclude that E divides H in $\mathbb{F}[X, Y]$. This means that we can find a polynomial $K(X, Y)$ such that $H(X, Y) = E(X, Y)K(X, Y)$.

Simple counting of degrees shows that the degree of K is $\leq (K_X, K_Y)$. Additionally, for every x_i in our set, we have

$$H(x_i, Y) = Q_{i,Y}(Y)E(x_i, Y) = K(x_i, Y)E(x_i, Y)$$

and thus provided $E(x_i, Y) \neq 0$, $Q_{i,Y}(Y) = K(x_i, Y)$. Identical logic holds for each y_j. □

A.2 Low Degree Replacement

In this subsection, we combine the above results on bivariate polynomials to demonstrate the main claim that we cite in the proof of Theorem 3.2. The idea is that, given two nearby bivariate polynomials $F, G \in \mathbb{F}[X, Y]$, where each is low degree in a distinct variable, we can construct a third polynomial $K \in \mathbb{F}[X, Y]$ that is close to both F and G and low-degree in both variables.

Theorem A.1. *Let $F(X, Y)$ and $G(X, Y)$ be bivariate polynomials on domain $\mathbf{X} \times \mathbf{Y}$ with degrees*

$$\deg_X F \leq M \qquad \deg_Y F \leq n$$
$$\deg_X G \leq m \qquad \deg_Y G \leq N$$

such that $M > m$ and $N > n$. Furthermore, suppose the relative Hamming distance between F and G is bounded:

$$d(F, G) < \sigma^* \tau^*,$$

where

$$1 > \sigma^* + \tau^* + \frac{m}{|\mathbf{X}|} + \frac{n}{|\mathbf{Y}|}. \tag{43}$$

Then there exists a polynomial $K(X,Y)$ with the following properties:

1. *It is low degree in both variables:*

$$\deg_X K \le m \qquad \deg_Y K \le n$$

2. *It is close to both F and G as univariate specializations:*

$$d(K(-,Y), F(-,Y)) \le \sigma^*$$
$$d(K(X,-), G(X,-)) \le \tau^*$$

Proof. Consider the difference set

$$S = \{(x,y) \in \mathbf{X} \times \mathbf{Y} \mid F(x,y) \neq G(x,y)\}$$

By Lemma A.2, there exists nonzero $E(X,Y)$ with degree

$$\deg_X E \le \sigma^*|\mathbf{X}| \qquad \deg_Y E \le \tau^*|\mathbf{Y}|$$

such that $E|_S = 0$ and hence for all $(x,y) \in \mathbf{X} \times \mathbf{Y}$,

$$E(x,y)F(x,y) = E(x,y)G(x,y).$$

Note that these have degrees

$$\deg_X EF \le M + \sigma^*|\mathbf{X}| \qquad \deg_Y EF \le n + \tau^*|\mathbf{Y}|$$
$$\deg_X EG \le m + \sigma^*|\mathbf{X}| \qquad \deg_Y EG \le N + \tau^*|\mathbf{Y}|$$

By Proposition A.1, there exists $H(X,Y)$ satisfying

$$H(x,y) = E(x,y)F(x,y) = E(x,y)G(x,y) \tag{44}$$

for all $(x,y) \in \mathbf{X} \times \mathbf{Y}$, and with degree

$$\deg_X H \le m + \sigma^*|\mathbf{X}| \qquad \deg_Y H \le n + \tau^*|\mathbf{Y}|$$

Fix an $x_0 \in \mathbf{X}$, and note that by (44), $H(x_0,y) = E(x_0,y)F(x_0,y)$ for all $y \in \mathbf{Y}$ and thus that $\deg_Y H = \deg_Y EF \le n + \tau^*|\mathbf{Y}|$. Observe that, (43) can be weakened and rearranged to

$$n + \tau^*|\mathbf{Y}| < |\mathbf{Y}| \tag{45}$$

This implies the equality of polynomials

$$H(x_0,Y) = E(x_0,Y)F(x_0,Y)$$

Thus, by Lemma A.4, $E(X,Y)|H(X,Y)$ as polynomials in $\mathbb{F}[X,Y]$. Define

$$K(X,Y) = \frac{H(X,Y)}{E(X,Y)},$$

which we now show satisfies the desired properties.

1. The degree of K is computed as follows.

$$\deg_X K = \deg_X H - \deg_X E \leq m$$
$$\deg_Y K = \deg_Y H - \deg_Y E \leq n$$

2. As the other will follow from symmetry, we will prove just the first inequality. In particular, we demonstrate the following chain of inequalities.

$$d(K(-,Y), F(-,Y)) = \mathbb{P}_x[K(x,Y) \neq F(x,Y)] \tag{46}$$
$$\leq \mathbb{P}_x[E(x,Y) = 0] \tag{47}$$
$$\leq \sigma^* \tag{48}$$

   To show (47), we prove the contrapositive

$$E(x,Y) \neq 0 \in \mathbb{F}[Y] \implies K(x,Y) = F(x,Y) \in \mathbb{F}[Y].$$

   Let $x \in \mathbf{X}$ be such that $E(x,Y) \neq 0$. Since $\deg_Y E \leq \tau^*|\mathbf{Y}|$, the polynomial $E(x,Y)$ can have at most $\tau^*|\mathbf{Y}|$ zeros, and hence—since K and F must be equal when E is nonzero—the polynomials must agree on at least $(1 - \tau^*)|\mathbf{Y}|$ points. Since $\deg_Y K = \deg_Y F \leq n$, these are forced to be equal as polynomials in $\mathbb{F}[Y]$ as long as $n < (1 - \tau^*)|\mathbf{Y}|$, which follows from rearranging (45).

   To prove (48), we equivalently show that $\mathbb{P}_x[E(x,Y) \neq 0] \geq 1 - \sigma^*$. We simply argue that for $E(x,Y) \neq 0$ in $\mathbb{F}[Y]$, it suffices that there is at least one coefficient—given as a polynomial in x—that is nonzero. Since $\deg_X E \leq \sigma^*|\mathbf{X}|$, any such coefficient is nonzero on at least $(1 - \sigma^*)|\mathbf{X}|$ points, and hence $\mathbb{P}_x[E(x,Y) \neq 0] \geq 1 - \sigma^*$.

□

References

[1] Ben-Sasson, E., Carmon, D., Ishai, Y., Kopparty, S., and Saraf, S. Proximity gaps for reed–solomon codes. *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)* (2020), 900–909.

[2] Spielman, D. Computationally efficient error-correcting codes and holographic proofs.